

**ES234422 – PEMODELAN & ANALITIKA PREDIKTIF**  
**SEMESTER GENAP 2023/2024**  
**TUGAS GROUP PROJECT #2 (TGP #2)**  
**KLASIFIKASI**

Tugas kelompok kedua merupakan kelanjutan dari TGP #1 dan terkait dengan topik klasifikasi. Tugas ini untuk memberikan pengalaman kepada mahasiswa dalam mengimplementasikan berbagai metode klasifikasi menggunakan python. Data yang digunakan sama dengan data untuk TGP #1, yaitu data mengenai kampanye pemasaran langsung (*direct marketing*) produk deposito bank berjangka dari sebuah institusi perbankan di Portugal. Kampanye dilakukan melalui panggilan telpon di mana seorang klien mungkin harus dihubungi lebih dari sekali. Kampanye melalui pemasaran langsung ini ditujukan untuk **memprediksi apakah seorang klien yang dihubungi akan membeli produk deposito berjangka tersebut atau tidak**.

**A. Deskripsi data**

Data kampanye pemasaran langsung adalah data yang diperoleh dari bulan Mei 2008 hingga November 2010. Data ini terdiri dari 45.211 baris, 16 atribut (variabel independen) dan 1 atribut target (variabel dependen).

**# bank client data:**

- (1) **age** (numeric)
- (2) **job**: type of job (categorical:  
"admin.", "unknown", "unemployed", "management", "housemaid", "entrepreneur", "student", "blue-collar", "self-employed", "retired", "technician", "services")
- (3) **marital**: marital status (categorical: "married", "divorced", "single"; note: "divorced" means divorced or widowed)
- (4) **education** (categorical: "unknown", "secondary", "primary", "tertiary")
- (5) **default**: has credit in default? (binary: "yes", "no")
- (6) **balance**: average yearly balance, in euros (numeric)
- (7) **housing**: has housing loan? (binary: "yes", "no")
- (8) **loan**: has personal loan? (binary: "yes", "no")

**# related with the last contact of the current campaign:**

- (9) **contact**: contact communication type (categorical: "unknown", "telephone", "cellular")
- (10) **day**: last contact day of the month (numeric)
- (11) **month**: last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec")
- (12) **duration**: last contact duration, in seconds (numeric)

**# other attributes:**

- (13) **campaign**: number of contacts performed during this campaign and for this client (numeric, includes last contact)
- (14) **pdays**: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)
- (15) **previous**: number of contacts performed before this campaign and for this client (numeric)
- (16) **poutcome**: outcome of the previous marketing campaign (categorical: "unknown", "other", "failure", "success")

**Output variable (desired target):**

- (17) **subscribe**: has the client subscribed a term deposit? (binary: 'yes', 'no')

**B. Tugas**

1. Lakukan perbaikan eksplorasi data dan juga praproses dari TGP #1 (jika diperlukan) untuk keperluan implementasi model prediksi (klasifikasi). Lakukan pendalaman pemahaman terhadap proses bisnis dari studi kasus ini dari berbagai sumber, sehingga perbaikan eksplorasi dan praproses data menjadi lebih tepat sasaran.

2. Lakukan proses prediksi (klasifikasi) dengan berdasarkan pada variabel target (variabel dependen **subscribe**). Lakukan implementasi menggunakan berbagai metode: (i) **decision tree**, (ii) **k-NN**, (iii) **Naïve Bayes**, (iv) **SVM**, (v) **BPNN**, (v) **Logistic Regression**, dan (vi) **Random Forest**.
  - a. Untuk setiap metode gunakan *library* python yang anda pahami dan menurut anda paling baik, termasuk *library* yang diberikan dalam sesi tutorial. Jelaskan penentuan parameter (*parameter setting*) *library* yang anda gunakan.
  - b. Untuk pembangunan model kasifikasi, lakukan pembagian keseluruhan data menjadi data pelatihan (*training*) sebesar 70% dan data uji (*testing*) 30%. Data pelatihan digunakan untuk membangun dan memvalidasi model klasifikasi, sedang data uji digunakan sebagai data pengujian final menggunakan model terbaik.
  - c. Untuk membangun model terbaik, gunakan metode **10-cross validation** pada data pelatihan yang diulang 3 (tiga kali). Hitung nilai **f1-measure** dan juga **AUC-ROC** untuk melakukan evaluasi kinerja dari model klasifikasi untuk setiap metode yang diminta.
  - d. Untuk masing-masing metode, hitung nilai rata-rata dan juga simpangan baku dari hasil model klasifikasi menggunakan **10-cross validation**. Lalu pilih hasil terbaik sebagai model terpilih untuk setiap metode klasifikasi.
  - e. Anda diberi kebebasan menentukan parameter yang akan digunakan dalam eksperimen untuk membangun model klasifikasi terbaik pada setiap metode klasifikasi. Tuliskan referensi yang digunakan beserta penjelasan singkat mengenai parameter yang digunakan.
  - f. Dengan menggunakan model klasifikasi terbaik untuk masing-masing metode klasifikasi, hitung kinerja dari masing-masing model dan bandingkan hasilnya menggunakan data uji.
  - g. Dari model klasifikasi terbaik (untuk setiap metode kalsifikasi) dari hasil perbandingan langkah (f), lakukan eksperimen dampak dari seleksi fitur (reduksi dimensi/kolom/fitur data). Untuk ini lakukan reduksi fitur mulai dari total 16 fitur variable independent hingga hanya tersisa 10 fitur saja (lakukan eksperimen jumlah fitur terbaik antara 10 s.d. 16). Gunakan metode **backward selection** dengan menggunakan *library* untuk seleksi fitur (anda bebas mencari sendiri *library* tersebut dan berikan uraian singkat mengenai *library* yang dipakai). **Backward selection** diawali dengan melatih model menggunakan semua fitur dan lakukan evaluasi kienerja model menggunakan *f1-measure*. Kemudian pilih satu fitur dari semua fitur yang ada untuk dihapus, yaitu hapus fitur yang memberikan evaluasi kinerja sama dengan atau lebih baik dari sebelum penghapusan. Dengan cara yang sama lakukan seleksi sebuah fitur yang akan dihapus berikutnya hingga tercapai sisa target fitur yang diharapkan.

### C. Laporan dan Batas Waktu

- Laporan ditulis pada kertas berukuran A4 dengan spasi tunggal. Laporan dalam format PDF diserahkan per kelompok dan diunggah melalui myITS Calsroom **paling lambat pada tanggal 5 Mei 2024 pukul 23:59 WIB** (hanya satu orang dari setiap kelompok yang dapat menyerahkan laporan).
- Penilaian akan didasarkan pada aspek: sistematika penulisan dan kelengkapan laporan (10%), kejelasan uraian laporan termasuk tata tulis (10%), kejelasan metode dan uraian hasil eksplorasi dan praproses data (15%), kejelasan metode pembangunan dan pengujian model klasifikasi (40%), dan presentasi melalui zoom meeting yang melibatkan semua anggota kelompok (25%).
- Selain laporan tugas, setiap kelompok juga harus mengunggah *script program python (saved file)* dalam *Jupiter Notebook* atau *Google Colab Notebook*
- Setiap kelompok harus mempresentasikan hasil implementasi menggunakan zoom meeting.
- Kedua dokumen dari TGP #2 (File PDF dan Script Python) dan file video rekaman presentasi zoom meeting (dalam fiirmat MP4) dijadikan satu file dalam format **.zip** atau **.7z** atau **.rar** (nama file: **TGP2\_nomorkelompok**)
- Kalau ada pertanyaan sehubungan dengan TGP #2 silakan hubungi saya melalui WA.
- **Isi laporan dan script program yang mengindikasikan adanya plagiarisme tidak akan dinilai.**

-----oooOooo-----